METHOD AND APPARATUS FOR CLASSIFYING UNMARKED STRING SUBSTRUCTURES USING MARKOV MODELS

Abstract of the Disclosure

A technique for structurally classifying substructures of at least one unmarked string utilizing at least one training data set with inserted markers identifying labeled substructures. A model of class labels and substructures within strings of the training data set is first constructed. Markers are then inserted into the unmarked string, identifying substructures similar to substructures within strings of the training data set by using the model. Finally, class labels of the substructures in the unmarked string similar to substructures within strings of the training data set are predicted using the model.

)